JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Predicting Depression Using Machine Learning Algorithms

Ishfaq Majeed [a], Dr. Simmi Dutta [b], Mr. Prabjot Singh [c].
Department of Computer Science and Engineering, Govt. College of engineering and technology, Jammu, [a]
H.O.D Department of Computer Science and Engineering, Govt. College of engineering and technology, Jammu [b]
Astt.prop Department of Computer Science and Engineering, Govt. College of engineering and technology, Jammu [c]
Jammu and Kashmir, India

Abstract

Depression is a serious psychological health problem that can disrupt a person's emotional well-being. In this contemporary world mental stress has emerged as a significant concern, particularly affecting students who were once perceived as carefree. Escalating stress levels among students have been linked to many issues like depression, suicide, heart attacks, and strokes. Detecting students' depression early is very important nowadays which is the primary objective of this research paper. This research focuses on evaluating depression experienced by college students, using the depression scale questionnaires (PHQ 9). Data was collected from 300 students of the GCET Jammu. They were questioned simply about how they felt in situations they might have encountered within the previous two weeks. Their answers are given some amount of weight that helps to calculate a score to analyse the depression level of students. There are five machine learning algorithms (MLA) that are used: Support Vector Machine (SVM), Decision tree (DT), k-nearest neighbour (KNN), Random Forest (RF), and Naïve Bayes (NB). We have also utilized performance evaluation with a confusion matrix in this work. The results show that, when compared to other algorithms, SVM provides the highest accuracy of 94.8%.

Keywords: Depression, Machine Learning, Performance Evaluation.

1. Introduction.

Depression and Anxiety are among the leading causes of disability, affecting an estimated around 245 million people worldwide [20]. According to the recent Lancet Committee report, mental health disorders will become more prevalent across the board and will have a negative economic impact on the global economy of trillions of dollars per year owing to fretfulness and depression alone by 2030[12-13]. In 2018, the World Health Organization published guidelines on how to treat the physical troubles that individuals with severe mental disorders may encounter, which if disregarded, could result in a substantial lifetime burden of the condition. [20, 12]. An extreme negative emotion, such as melancholy, pessimism, or despondency, to the point where it interferes with day-to-day activities, is a defining feature of the mental health disease known as depression. It is like a unique kind of illness that affects how we think and feel inside. Sometimes people also call it Major Depressive Disorder. It has become a severe issue in the present generation. Several factors have contributed to the daily increase in the overall number of cases like pressure at college, work, personal life, etc. Lots of people around the world, nearly 300 million have this problem [11]. When someone has depression from a medical perspective, can have various symptoms such as difficulty in remembering things, finding it hard to focus or pay attention, Struggling to make decisions, mood going up and down a lot, feeling guilty, suicidal thoughts, etc[2]. Depression is the leading reason why several people choose to end their lives, accounting for a maximum (more than two-thirds) of the suicides that take place each year [15]. Every hour, a student from different parts of our country takes their own life. Many students in our nation between the ages of 14 and 30 commit suicide, according to a 2012 Lancet investigation. We can use this information to understand how stressed students are feeling early on [16]. If we can determine how stressed they are, we can either help them recover swiftly or gradually. De Choudhury et al. [17] found that they could determine whether a person has a major depressive condition or if they could experience depression in the future based on their social media activity and tweets. Depression tends to slow down the daily routine performances due to the absence of attention and curiosity. Ultimately, it harms both the body and the brain of an individual if it remains undetected. More people should be educated about depression so that those who are affected can receive the support they need. With proper support and treatment, people can get back to feeling better and living their lives more healthily. There are currently about 4.3 billion social media users globally as a result of technology improvements. [3]. Social media has grown in importance for people. They have quick ways of expressing their thoughts, feelings, and opinions. Individuals use social media as a platform to vent their frustrations, which helps them feel better. Through sharing the thoughts that would otherwise keep bothering them in their heads, social media has become a platform for people to express their innermost feelings and sentiments. Different machine learning algorithms have been used by different researchers to predict psychological disorders (depression), and it has been found that the performance of each algorithm varies depending on the social media data, making it impossible to identify a single algorithm that is always the most effective. Only a small number of researchers have so far modified machine learning methods to recognize and forecast depression. This research has presented a ground-breaking solution to close this gap. Different machine learning methods, correlation, and performance evaluation have all been used.

2. Related work.

This kind of research problem has incredibly intricate components that need careful examination. This section examined several relevant research publications to ascertain the methodologies and approaches used in the earlier works and the research gaps. Recent technological breakthroughs and advancements in Artificial intelligence/Machine Learning approaches have enabled the development of more effective prediction and decision-making tools. To appreciate how machine learning helps diagnose mental depression, many

research papers and articles have been written and shared in various publications. Two main kinds of study are commonly used to identify depression through social media communications: one focuses on the text's substance, while the other examines the traits or descriptions of the user who is posting the messages. ***Md. Rafiqul Islam et al [1***] proposed a machine learning-based technique for depression analysis using Facebook data. To obtain the necessary data for the investigation, the scientists employed the Ncapture technique. They investigate each aspect (emotional, temporal, and linguistic style) separately using various supervised machine-learning techniques. A Decision Tree (DT) gives the highest accuracy another ML approach to finding depression. The analytics performed on the selected dataset results in what depression is and what the common factors contributing to depression are. ***G. Geetha et al [2]*** proposed machine learning for the early detection of depression, using social platform data. In this study, various machine-learning algorithms were used and logistic regression achieved the highest accuracy. In this study, it was revealed that the decision of depression prediction makes sure whether the end user needs the medical treatment or not. ***Ravinder Ahuja et al [3]*** investigated the mental stress experienced by college students at different stages of their lives. They used a dataset of 206 students from the Jaypee Institute of Information Technology, collecting data through the Perceived Stress Scale test. Linear regression, Naïve Bayes, Random Forest, SVM, and Random Forest were among the classification methods used. SVM yielded the highest accuracy, around 85%, in predicting stress levels. ***Tadesse et al [5]*** Discussed how depression plays a major part in the global rates of disability and suicide. This study uses natural language processing (NLP) and machine learning techniques to detect depressive attitudes among Reddit users. By identifying a lexicon of depression-related terms, the study achieves significant performance improvements in depression detection. With integrated features (LIWC+LDA+bigram) and the Multilayer Perceptron (MLP) classifier, the best results are obtained with 91% accuracy and a 0.93 F1 score. ***B. Zohuri et al [6]*** Mood disorders are intimately linked to depression and are connected with a higher incidence of suicide thoughts. One must have persistent symptoms for several weeks to be diagnosed with depression, a mood disorder that has a major influence on day-to-day functioning. Many deaths each year, particularly in the Asia contenient, are caused by suicide, which is a serious public health issue. Preventive measures depend heavily on early diagnosis of depression and suicidal thoughts. To identify indicators of depression and the risk of suicide, this communication examines the possibilities and constraints of artificial intelligence (AI), particularly as it relates to machine learning and deep learning. It also emphasizes the significance of effective solutions to manage related difficulties and the effects that traumatic events like the COVID-19 pandemic have on mental health***. Chiong et al [8]*** findings confirm that depression is a major contributor to global suicides, often undiagnosed. Social media posts by individuals with depression can predict their condition. This study explores machine learning's effectiveness in identifying depression signs in social media text, even when explicit keywords are absent. Using training data from Twitter and testing data from non-Twitter, several text pre-processing and machine learning algorithms were evaluated. According to the results, the strategy is effective even in the absence of specific keywords and is applicable to a variety of social media platforms. This study adds to the body of knowledge about social media's usefulness as a tool for tracking mental health and emphasizes its potential for wider uses in the identification of depression on a variety of platforms. ***Deshpande et al [7]*** Depression is a prevalent mental health issue associated with an increased risk of early mortality, suicidal thoughts, and significant impairment in daily life. In the field of Emotion Artificial Intelligence, research on emotion detection, particularly in text mining, is ongoing. With the abundance of user-generated data on internet-based platforms, sentiment analysis of text and images has gained prominence. This study leverages Natural Language Processing to analyze Twitter feeds, focusing on depression. By classifying individual tweets as neutral or negative using a curated word list, support vector machine, and Naïve-Bayes classifiers, the research achieves commendable results in depression detection. Future work could involve expert-based input to enhance precision and reduce false positives in sentiment analysis for improved depression detection. ***Dr. S. Smys et al [11***] The necessity for early emotional state identification has grown in importance due to the extensive usage of social media and the internet. Millions of people are impacted by psychiatric diseases, which carry serious hazards. By highlighting the possibility of early detection to lessen the effect of various conditions, this study seeks to solve these difficulties. The researchers propose a machine learning method that combines Naïve Bayes and support vector machines to predict depression. Using a range of textual, semantic, and writing content elements, this hybrid approach evaluates numerous deep learning methods for early prediction.. According to the results, the hybrid algorithm performs better than single classifiers and achieves a greater prediction accuracy. The study highlights how early prediction could be improved by leveraging Twitter data to measure depression through online media posts and behaviors**.**

It is evident from the linked paper above that the bulk of studies conducted to now on depression prediction have frequently used information from social media sites such as Facebook and Twitter. However, as people frequently behave differently on social media than they do in real life, these datasets might not be very trustworthy. Many posts on social media are copied from others or taken from various sources, making it challenging to make an accurate diagnosis based solely on this information. Our work adopts a different strategy to address this problem. Instead of relying solely on social media data, we ask persons a series of standardized questions. Based on their responses to these questions our model, which has been trained on a dataset of such responses makes predictions about their mental health. We want to understand how stressed people are by asking them specific questions, like the Patient Health Questionnaire (PHQ-9), which is a standard set of questions. Aim to make the process of screening for depression more accurate and efficient. In order to accomplish this, we are examining respondents' answers to the PHQ-9 questions using machine learning methods. These algorithms can find patterns in the responses that might suggest someone is experiencing depression. Ultimately, we are working on creating a machine-learning model. This model will be good at figuring out if someone is dealing with depression based on how they answer the PHQ-9 questions. This way, we can help identify depression in individuals more accurately and quickly using this standard questionnaire.

3. Materials and Methods

This research focuses on the detection of depression among students using the depression scale questionnaire (PHQ 9).

A. 3.1 Dataset

The data was taken from 300 college students from GCET Jammu. They were asked straightforward questions regarding their feelings in hypothetical situations they might have encountered in the preceding two weeks.. Students' responses are assigned a certain number of weights, and these weights are used to compute a score that allows for the analysis of the students' depression levels. The classifiers read a CSV file that contains 15,000 records from the dataset. After the dataset was divided into 70:30 segments, which represent the training and test sets, respectively, machine learning techniques, such as Decision Tree, Random Forest Tree, Naïve Bayes, Support

Vector Machine, and KNN were used to classify it. The Python implementation of the machine learning techniques was done with Visual Studio. This suggests whether the student is depressed or not, depending on how severe the symptoms are.

3.2 Participants

A total of Three hundred students from various semesters participated in the study and answered questions about themselves. These were simple inquiries concerning the emotions that they may have encountered in the previous two weeks.

B. Methodology

The Depression Scale questionnaire, PHQ-9, was used to gather data for the study. The instrument used to assess students' levels of depression is the Patient Health Questionnaire (PHQ-9). The nine questions on the self-administered questionnaire measure the occurrence and intensity of depression symptoms throughout the last two weeks.

A wide range of symptoms that are frequently linked to depression are covered by the PHQ-9, such as depressive moods, lack of interest in activities, changes in eating or sleep patterns, and suicidal thoughts. Patients are asked to rate each item on a scale of 0 to 3, where 0 represents no symptom experience at all and 3 represents almost daily symptom experience for the previous two weeks. The final result is a total score out of 27, which is calculated by adding the scores for each item.

Scores of 0-4 indicate minimal or no depression.

5-9 indicate mild depression

.

10-14 indicate moderate depression.

15-19 indicate moderately severe depression.

20 or higher indicates severe depression.

Table 1. PHQ-9.

Questions	Not at all	Several days	Over half of the days	Nearly Every day
1. Low motivation or enjoyment in tasks.	0	1	2	3
2. Feeling down, depressed, or hopeless.	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much.	0	1	2	3
4. Feeling tired or having little energy.	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling inferior to others, believing you are a failure, or believing you have let your family or yourself down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Being so slow to talk or move that others would have observed. Or, on the other hand, being agitated or restless and moving around a lot more than normal	0	1	2	3
9. Thoughts that you would be better off dead, or hurting yourself	0	1	2	3

Table 2. Interpretation of the Overall Score

Total Score	Depression Severity
1-4	Minimal depression
5-9	Mild depression
10-14	Moderate depression
15-19	Moderately severe depression
20-27	Severe depression

C. Machine learning Algorithms

The model was trained with five machine-learning algorithms.

a. Decision tree.

The decision tree method in machine learning is like making a series of choices step by step, just like a tree with branches as depicted in Figure 2. It's really good for solving problems where you want to make predictions. Decision trees are simple to comprehend, and they stay pretty much the same each time you use them. They can be used for sorting things into different categories (like "yes" or "no") or for estimating values (like numbers).

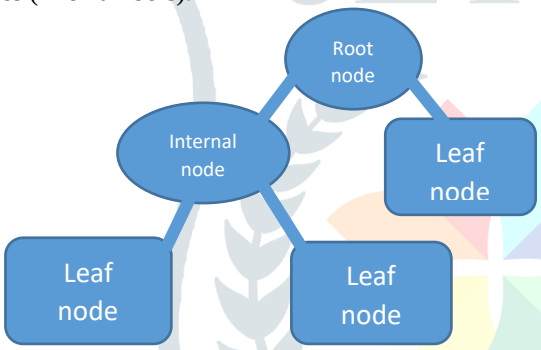


Figure 2 Structure of DT approach

b. *Support Vector Machine (SVM)*

SVM is a computer algorithm used in depression detection. It works by finding a clear line or boundary that best separates people with depression from those without. Think of it as drawing a line on a graph to separate two groups of points. The goal is to find the best line that maximizes the gap between the two groups, making it easier to tell if someone is depressed or not based on certain characteristics. SVMs are like smart detectives that help analyze data and make predictions about a person's mental health.

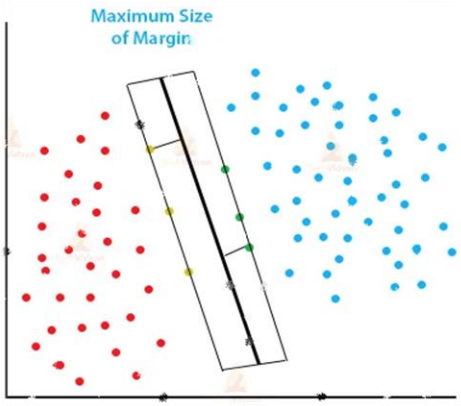


Fig 3 Support vector machine representation

c. Random Forest.

One effective machine-learning technique for identifying depression is called Random Forest. Consider it as a collection of numerous decision trees, representing various detectives examining the same set of data. After voicing their opinions, the trees collectively decide whether or not a person is likely to suffer from depression. When multiple detectives (or decision trees) work together, the forecast produced is frequently more accurate. Random Forests function similarly to a group of friends collaborating to estimate an individual's mental health by taking into account a variety of behavioural and linguistic cues.

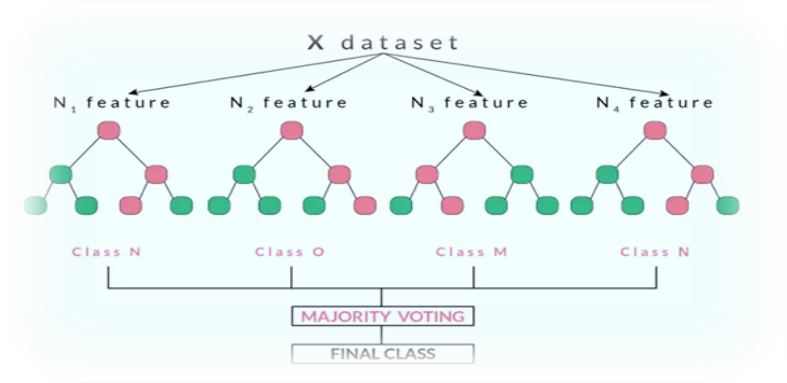


Figure 4 Random Forest

d. Naive Bayes

Naive Bayes is a smart algorithm for detecting depression. It is like a detective that uses probability to make guesses. Naïve Bayes looks at how people with depression typically answer phq-9 questions. It assumes that each question's answer is independent, which is why it's called "naïve." By comparing your answers to patterns it has learned from others, Naïve Bayes estimates the likelihood of you having depression. It is a quick and helpful way to identify potential depression based on your responses to the PHQ-9 questions. The formula for Bayes theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where.

P (A|B) is Subsequent probability: Probability of suggestion A on the experiential event B.

P (B|A) is Likelihood probability: The probability of the evidence given that the probability of a hypothesis is true.

P (A) is Prior Probability: The probability of suggestion before observing the evidence.

P (B) is Marginal Probability: Probability of Evidence.

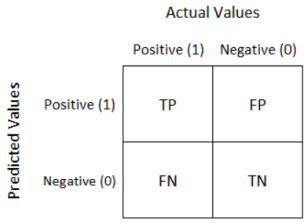
e. K-Nearest Neighbour

Based on the Supervised Learning method, K-Nearest Neighbour is one of the most basic machine learning algorithms. The K-NN algorithm classifies the new case in the category most comparable to the existing categories based on its assumption that the new instance and its data are similar to the previous cases. It is usually used for Classification difficulties, although it may also be used for Regression and Classification.

K-Nearest Neighbours is a straightforward effective machine learning algorithm often used in tasks like depression detection. Think of it as a friendly neighbour who helps you decide if someone is depressed or not based on the people they live closest to. It's a friendly and straightforward way to estimate depression risk based on your PHQ-9 answers and how they compare to others.

4. Experiments and Results

We have gathered an actual data set for various techniques to identify depression for this paper. Like K-Nearest Neighbor (KNN), Naïve Bayes (NB), Decision Tree (DT), Random Forest Tree (RFT), and Support Vector Machine (SVM). These five models were trained, and their results varied when put to the test. The results are shown below using something called confusion matrices. In these matrices, the rows represent the actual classes, and the columns represent what our models predicted. The numbers 0, 1,2, 3, and 4, in both rows and columns, stand for different levels of severity (like normal, mild, moderate-severe, and severe cases respectively). An essential machine learning technique for assessing a classification model's performance is a confusion matrix. It offers a concise synopsis of the model's prediction performance. as we have demonstrated in Figures 4 through 8 of the Performance Evaluation.



- **True Positive (TP):** When the model correctly predicts something as positive.
- **False Positive (FP):** When the model wrongly predicts something as positive.
- **True Negative (TN):** When the model correctly predicts something as negative.
- **False Negative (FN):** When the model wrongly predicts something as negative.

From these, we calculate:

- **Precision:** How many of the predicted positives were correct?

  Precision Score = TP / (FP + TP)

- **Recall:** How many of the actual positives were correctly predicted?

  Recall Score = TP / (FN + TP)

- **Accuracy:** Overall correctness of the model's predictions.

  Accuracy Score = (TP + TN)/ (TP + FN + TN + FP)

- **Fi Score:** A balance between precision and recall.

  F1 Score = 2* Precision Score * Recall Score/ (Precision Score + Recall Score)

A "Classification Report" summarizes all of these metrics, providing insights into how well the model is performing.

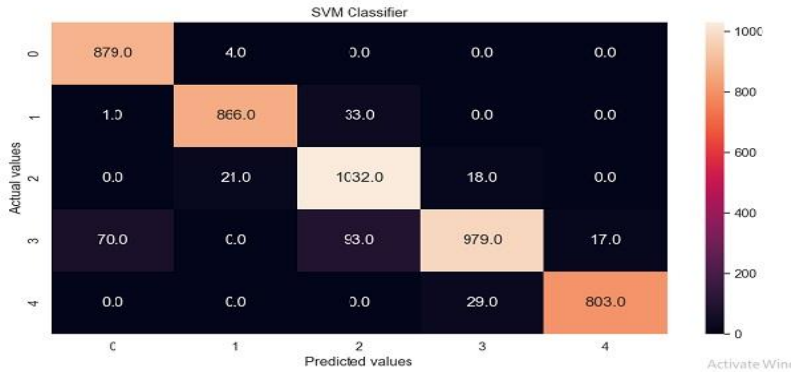


Figure 4. ***SVM classifier***

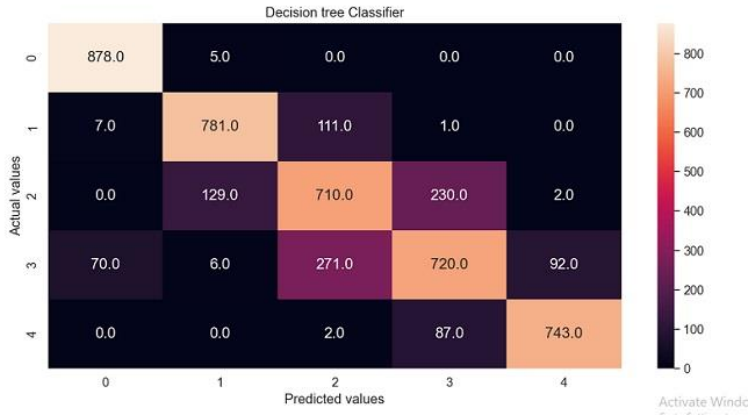


Figure 5. Decision tree

Figure 6. Random forest classifier

Figure 7. Naïve Bayes classifier

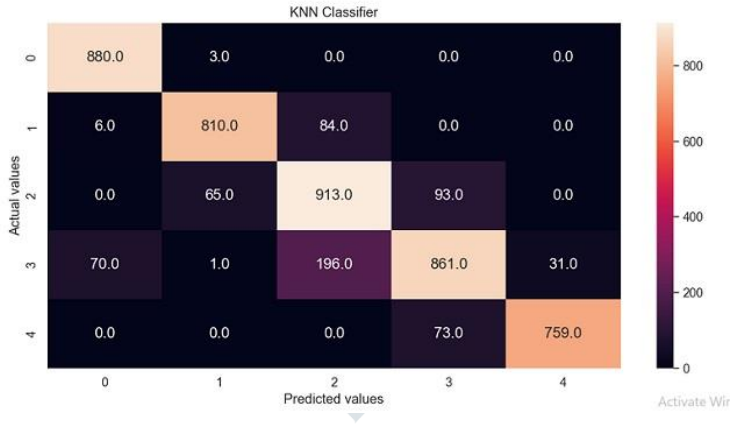
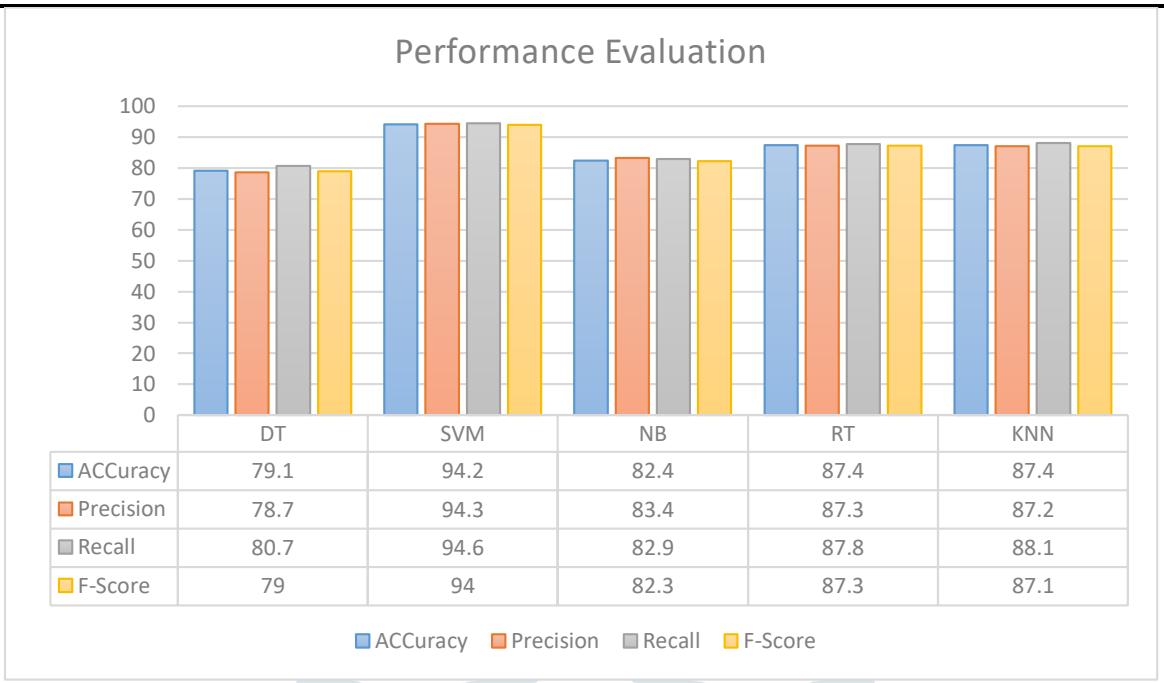


Figure 8. KNN classifier

The results obtained by different algorithms are given in Table below and its graphical representation is given in Fig below

Algorithm	*Accuracy*
SVM algorithm	*94*
Decision tree	*79*
Random forest algorithm	*87*
Naïve Bayes algorithm	*83*
KNN algorithm	*87*



	DT	SVM	NB	RT	KNN
ACCuracy	79.1	94.2	82.4	87.4	87.4
Precision	78.7	94.3	83.4	87.3	87.2
Recall	80.7	94.6	82.9	87.8	88.1
F-Score	79	94	82.3	87.3	87.1

5. Conclusion and future work.

Depression is a widespread global mental health issue and it is crucial to address it promptly. To address this issue effectively, early detection and increased awareness are important. Identifying depression early can lead to better treatment outcomes and even save lives. This work was initiated to help students/people by predicting depression in its early stages, allowing individuals to receive timely care. Extensive data analysis was conducted to gain a deep understanding of participants' behavior based on their responses to the Patient Health Questionnaire-9 (PHQ-9). This questionnaire captures information related to depression symptoms, providing valuable input for the predictive model. Based on the conclusions drawn from the trained model, the research results were categorized into five labels, representing different levels of depression severity. These labels help in identifying the varying degrees of the condition among individuals, providing a better understanding of the problem.

The research utilized a range of machine learning algorithms, including the Naïve-Bayes theorem, Support Vector Machines (SVM), k-nearest Neighbour (KNN), Decision Trees, and Random Forests. To assess how well each of these algorithms predicted depression, it was applied to the dataset. These algorithms' corresponding accuracy values are as follows, arranged in ascending order: Naïve Bayes classifier: 82.8% accuracy rate; Decision Tree classifier: 79.8.2% accuracy rate; Random Forest: 87.4% accuracy rate; KNN classifier: 87.5% accuracy rate; SVM: 94.1% accuracy rate. These accuracy values demonstrate how well these machine-learning methods work for early depression prediction. We intend to employ additional datasets in further work to confirm the effectiveness and efficiency of our methods for it to be reliable and provide a more encouraging outcome. Neural network-based models can also be built as an improvement to the present work to check their performance and precision

References

1. Md. Rafiqul Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M. Kamal, Hua Wang, and Anwaar Ulhaq, “Depression detection from social network data using machine learning techniques”,2018.
2. G. Geetha, R. Parthasarathy, and K. Thangadurai, "A Machine Learning Based Early Detection System for Depression using Social Platform Data," International Journal of Engineering and Technology, vol. 8, no. 3, pp. 1729-1734, 2016
3. Ravinder Ahuja, Alisha Banga, “Mental Stress Detection in University Students using Machine Learning Algorithms”,2019.
4. Anu Priya, K. Jeyalatha, and K. L. Shunmuganathan, "Prediction of Depression and Stress in Individual Life through Machine Learning Algorithms," International Journal of Applied Engineering Research, vol. 10, no. 15, pp. 35407-35411, 2015.
5. Tadesse, Michael M., Hongfei Lin, Bo Xu, and Liang Yang. "Detection of depression-related posts in Reddit social media forum." *Ieee Access* 7 (2019): 44883-44893.
6. Zohuri, Bahman, and Siamak Zadeh. "The utility of artificial intelligence for mood analysis, depression detection, and suicide risk management." *Journal of Health Science* 8, no. 2 (2020): 67-73.
7. Deshpande, Mandar, and Vignesh Rao. "Depression detection using emotion artificial intelligence." In *2017 international conference on intelligent sustainable systems (iciss)*, pp. 858-862. IEEE, 2017.
8. Chiong, Raymond, Gregorius Satia Budhi, Sandeep Dhakal, and Fabian Chiong. "A textual-based featuring approach for depression detection using machine learning classifiers and social media texts." *Computers in Biology and Medicine* 135 (2021): 104499.
9. Tuka Alhanai, Mohammad Ghassemi, James Glass, and Roger K. Pitman, "Detecting Depression with Audio and Textual Behavioral Signals," in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, ser. UbiComp '16, pp. 886-891, 2016.
10. Zhichao Peng, Tianqi Liu, Yuan Zhou, and Mingxuan Wang, "Multi-Kernel SVM Based Model for Recognizing Depressed People Using Microblog from Social Media," Mathematical Problems in Engineering, vol. 2016, Article ID 3048357, 10 pages, 2016.
11. Dr. S. Smys, K. Thangadurai, and A. Amudhavel, "Early Prediction of Depression from Social Media Content using Hybrid Machine Learning Techniques," International Journal of Control Theory and Applications, vol. 9, no. 44, pp. 149-159, 2016.

12. F. Bertini, D. Allevi, G. Lutero, D. Montesi, L. Calzà, Automatic speech classifier for mild cognitive impairment and early dementia, ACM Trans. Comput. Healthc. 3 (1) (2022) 1–11, http://dx.doi.org/10.1145/3469089.
13. R.C. Kessler, et al., Individual and societal effects of mental disorders on earnings in the United States: Results from the national comorbidity survey replication, Am. J. Psychiatry 165 (6) (2008) 703–711, http://dx.doi.org/10.1176/appi.app.2008.08010126.
14. Ding, Y., Chen, X., Fu, Q., & Zhong, S. (2020). A Depression Recognition Method for College Students Using Deep Integrated Support Vector Algorithm. International Journal of Environmental Research and Public Health, 17(17), 6249. doi: 10.3390/ijerph17176249.
15. Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt, "Detecting depression and mental illness on social media: an integrative review", 2017.
16. Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, Diana Inkpen, "Deep Learning for Depression Detection of Twitter Users",2018.
17. Priya A, Garg S, Tigga NP (2020) Predicting anxiety, depression and stress in modern life using machine learning algorithms. Procedia Computer Science 167:1258-1267.
18. .De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. Predicting depression via social media. ICWSM.13, 1-10 (2013).
19. Arias, Daniel, Shekhar Saxena, and Stéphane Verguet. "Quantifying the global burden of mental disorders and their economic value." *EClinicalMedicine* 54 (2022).
20. N. Votruba, G. Thornicroft, Global mental health, 2022, http://dx.doi.org/10.1017/gmh.2016.20
21. Srimadhur NS, Lalitha S (2020) An End-to-End Model for Detection and Assessment of Depression Levels using Speech. Procedia Computer Science 171:12-21.